Emerging Pandemic Threats PREDICT Project - Coronavirus Extension

The PREDICT Consortium strengthened global preparedness for emerging threats, in particular to detect viruses that may have the potential to spillover from wild animal hosts to people. As part of USAID's Emerging Pandemic Threats program, PREDICT was implemented from October 2009 through September 2020 by the PREDICT Consortium, as a multi-institutional cross-disciplinary team with numerous global, implementing and government partners in 30 countries (see https://ohi.vetmed.ucdavis.edu/programs-projects/predict-project/authorship for a list of contributors). This project pioneered a One Health approach to emerging infectious virus surveillance and risk communication at high-risk human-wildlife interfaces. This data asset and related datasets contain the test results of the PREDICT Coronavirus Extension Project, implemented from March 2020 through September 2020 to perform additional coronavirus PCR testing on archived animal specimens collected from 2009 to 2019.

This data asset includes two datasets from PREDICT describing the Coronavirus PCR test results from wildlife sampled in Asia, including Cambodia, Laos, Malaysia, Myanmar, Nepal, Thailand, and Vietnam. Tests results represent current findings of the PREDICT Coronavirus Extension Project, however, additional work to further characterize and confirm some results is ongoing. More details on the PREDICT project can be found at: https://p2.predict.global/

Dataset Contents

The PREDICT Emerging Pandemic Threats Project - Coronavirus Extension Data Asset consists of two datasets. Details on data field contents can be found in the Data Dictionary.

1. **PREDICTPCRTestsCoronavirus_P2**

   Each record describes the PCR tests conducted and the test results. Multiple tests were conducted for each individual sampled. The specimens tested were collected from 2014 to 2019 and originated from the PREDICT project and the archives of partners/collaborators. Additional information about some of these samples may be found in other PREDICT datasets within USAID's Data Development Library; see Data Asset "PREDICT Emerging Pandemic Threats Project" for more information.

2. **PREDICTPCRTestsCoronavirus_P1**

   Each record describes the PCR tests conducted and the test results. Multiple tests were conducted for each individual sampled. The archived specimens tested were collected from 2009 to 2014 and originated from the PREDICT project.

Related Data Assets & Datasets

Data Asset: PREDICT Emerging Pandemic Threats Project

Datasets: PREDICTSite/EventCharacterization (including event module datasets), PREDICTAnimalsSampled

Considerations and Limitations

The sites chosen and animals sampled were selected intentionally to accomplish project goals, which may have differed by country or had locality-based nuances, driven by potential viral presence, past outbreaks, common types of human-animal interactions, drivers for transmission and emergence specific to the sampled community, and local partner resources in those regions. As such, care should be taken in how demographics are interpreted and in how comparisons across locations are made. Data collected may not be representative of the broader population.

Rodents, shrews, and bats can be difficult to identify correctly in the field, and we took efforts to confirm species identity through genetic barcoding when possible. Analyses including these taxa should assess the column IDCertainty in AnimalsSampled to establish reliability of species identification (separate Data Asset: PREDICT Emerging Pandemic Threats Project). Genus or family level groupings may be necessary when few individuals of interest were barcoded.

Limitations of molecular detection: We employed consensus conventional polymerase chain reaction (PCR), which is a broadly reactive assay designed to identify all members (known and new) of a given viral genus and/or family. Some unknown viruses may fall outside of the detection limit of this test. PCR detects nucleic acid only and does not confirm the presence of infectious particles. A negative result only indicates that virus is not detectable by the method; a negative result does not confirm the absence of infectious particles. Results reported here are primarily short fragments of conserved genes that have been confirmed to be positive by sequencing. Real-time PCR was also conducted with conventional PCR and direct sequencing used to confirm results.

Definition of a new virus: A virus is considered to be new if it has equal or greater genetic variation than the difference between the two closest known virus species within a family/genus and if it represents a distinct (monophyletic) lineage. Our data are strongly suggestive of a new virus but such classification can only be conferred by the International Committee on Taxonomy of Viruses (ICTV).

Assessment of risk based on viral sequences: Results and their interpretation are provided on an individual test basis. Assessments of risk to public and animal health are based on what we know to be true for known viruses and relatedness of the viruses detected here to those known viruses.

Data Fields Included in each Dataset

See Data Dictionary attachment for full details.

Column Name	Column description
PREDICT_TestID	The numeric key to the test.
PREDICT_IndividualID	The numeric key to the individual tested. This can be used to link the P2 dataset to the AnimalsSampled dataset. Blank for samples that were not originally collected for PREDICT, but were tested from partners'/collaborators' archives as a part of the coronavirus extension project.
Country	The name of the country where the sampling occurred.
Latitude	The GPS location of the centroid of the site where the sample were collected in decimal degrees.
Longitude	The GPS location of the centroid of the site where the sample were collected in decimal degrees.
TaxaGroup	The taxonomic description of animal sampled.
SpeciesScientificName	The scientific name of the animal sampled. If complete genus species is unknown, this will be the next known level of taxonomic classification (i.e. genus, subfamily, family).
SampleDate	The date the animal was sampled in the field, if known.
SpecimenID	The text ID of the specimen used for the test.
SpecimenType	The type of specimen collected.
TestType	The type of PCR test performed.
ViralFamilyTested	The viral family that the PCR was performed for.
TestProtocol	The protocol of the PCR test performed and gene targeted.
TestResult	The result of the PCR test.
ConfirmationTestType	The type of test performed to confirm results.
ConfirmationResult	The confirmation test result.
Sequence	The virus sequence isolated. Will be blank for negative results.
GenbankAccessionNumber	The accession number to reference this sequence in Genbank. Will be blank for negative results.
Virus	The name of the virus detected in the sample. Will be blank for negative results.
ViralFamily	The taxonomic family the virus sequence belongs to. Will be blank for negative and Real-time PCR results.
Interpretation	Describes the zoonotic potential and threat of the detected virus to human health. Will be blank for negative results.